**UNIVERSIDAD PRIVADA SAN JUAN BAUTISTA**

**FACULTAD DE CIENCIAS DE LA SALUD**
**ESCUELA PROFESIONAL DE MEDICINA HUMANA**

**GRUPO 5 SEMANA 14**
**ASIGNATURA**
Sistematización y métodos estadísticos

**DOCENTE**
Castro Lopez Segundo Vicente

**INTEGRANTES**
Arianna Samantha Caballero Martinez
Lozano Laura Marina
Lizbeth Adriana Pachas Rojas
Nayely Luz Rojas Cortez
Alexander Manay Ventura

**CICLO:**
V

**LIMA-PERÚ**
**2025**

GRUPO 5

ALUMNOS:

- Arianna Samantha Caballero Martinez
- Lozano Laura Marina
- Lizbeth Adriana Pachas Rojas
- Nayely Luz Rojas Cortez
- Alexander Manay Ventura

## Instalar y cargar los paquetes

```{r}
install.packages("mice")
install.packages("ggmice")
```

```{r}
library(mice)
library(tidyverse)
library(here)
library(rio)
library(ggmice)
library(gtsummary)
```

# 1 Datos perdidos en estudios clínicos de cirrosis

En estudios clínicos sobre cirrosis hepática, es común encontrar datos faltantes en variables relevantes como niveles de enzimas hepáticas, cobre sérico o presencia de signos clínicos (ascitis, aracnoides, etc.). Estas ausencias pueden deberse a omisiones en el registro médico, limitaciones en la toma de muestras o pérdida de seguimiento del paciente.

En investigaciones biomédicas, eliminar los registros con datos faltantes el análisis de casos completos puede reducir el poder estadístico y sesgar los resultados

# 2 Imputación de datos

Para maximizar la información disponible y mejorar la calidad del análisis, aplicamos técnicas de imputación. En este análisis, emplearemos imputación múltiple para estimar los valores faltantes de forma robusta, especialmente en variables bioquímicas y clínicas, preservando la coherencia del conjunto de datos.

Este enfoque moderno permite mantener la muestra completa sin introducir sesgos sistemáticos, mejorando la validez de los resultados obtenidos en estudios de pacientes con cirrosis.

# 3 El dataset para este ejercicio

Para ilustrar el proceso de imputación múltiple de datos, utilizaremos el conjunto de datos `cirrosis`. Este dataset contiene información de 418 pacientes con diagnóstico de cirrosis hepática. Las variables registradas comprenden el estado del paciente (censurado, fallecido, trasplantado), días de seguimiento, edad, sexo, presencia de signos clínicos (ascitis, hepatomegalia, aracnoides, edema), y parámetros bioquímicos como bilirrubina, colesterol, albúmina, cobre, triglicéridos, enzimas hepáticas (SGOT, fosfatasa alcalina), recuento de plaquetas y tiempo de protrombina, entre otros.

Cargando los datos

```{r}
data_sm <- import(here("data", "cirrosis.csv"))
```

Un vistazo a los datos

```{r}
head(data_sm)
```

Un vistazo a los datos

```r
head(data_sm)
```

Description: df [6 × 20]

| | ID <int> | Dias_Segui... <int> | Estado <chr> | Medicamento <chr> | Edad <int> |
|---|---|---|---|---|---|
| 1 | 1 | 400 | Fallecido | D_penicilam... | 21464 |
| 2 | 2 | 4500 | Censurado | D_penicilam... | 20617 |
| 3 | 3 | 1012 | Fallecido | D_penicilam... | 25594 |
| 4 | 4 | 1925 | Fallecido | D_penicilam... | 19994 |
| 5 | 5 | 1504 | Censurado_... | Placebo | 13918 |
| 6 | 6 | 2503 | Fallecido | Placebo | 24201 |

6 rows | 1-6 of 20 columns

# 4 Realizando la imputación de datos

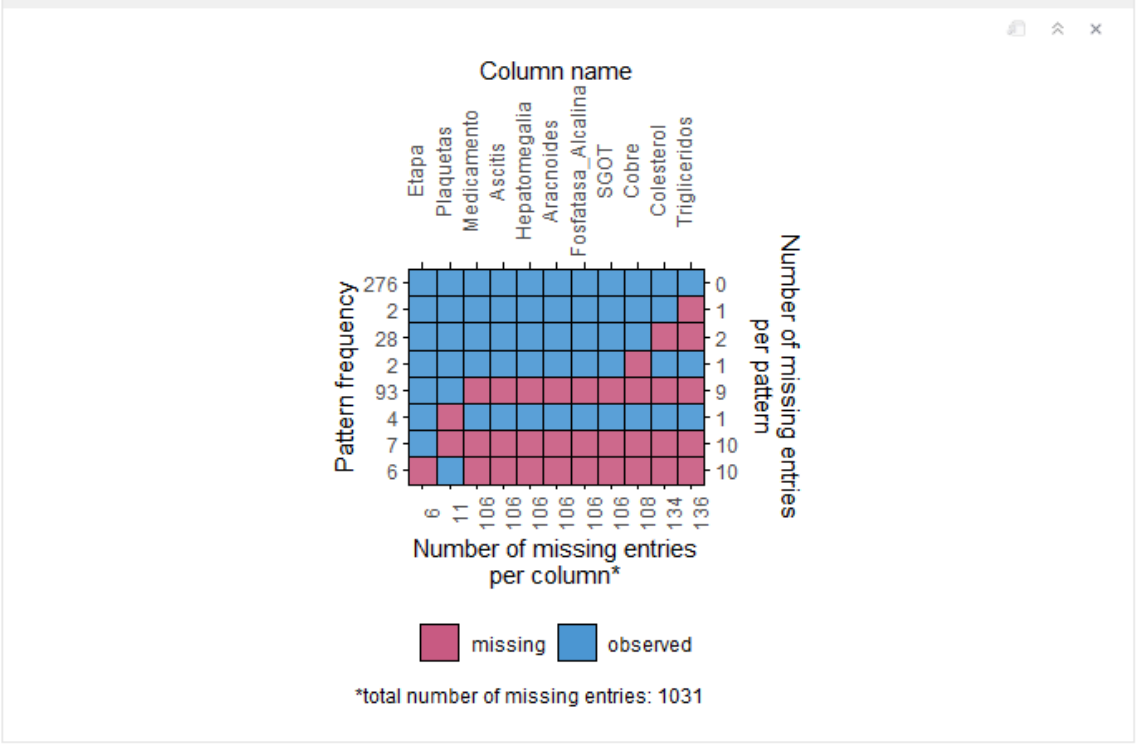
## 4.1 ¿Donde estan los valores perdidos?

Es importante saber en qué variables se encuentran los datos faltantes antes de iniciar la imputación. Una forma rápida es usando la función colSums() en combinación con is.na():

```r
colSums(is.na(data_sm))
```

También podemos visualizar los patrones de datos faltantes en forma de mapa de calor utilizando la función plot_pattern() del paquete ggmice. Aquí seleccionamos las variables con mayor proporción de valores faltantes para simplificar la visualización:

```r
data_sm |>
  select(
    Medicamento,
    Ascitis,
    Hepatomegalia,
    Aracnoides,
    Colesterol,
    Cobre,
    Fosfatasa_Alcalina,
    SGOT,
    Trigliceridos,
    Plaquetas,
    Etapa
  ) |>
  ggmice::plot_pattern(
    square = TRUE,
    rotate = TRUE
  )
```




La figura presentada muestra un mapa de calor que permite visualizar los patrones de datos faltantes en el conjunto de datos correspondiente al estudio de pacientes con cirrosis. En total, se identificaron 1031 valores perdidos distribuidos entre 11 variables clínicas y bioquímicas.

Se observa que varias variables comparten una cantidad idéntica de valores faltantes (106 casos), entre ellas: Medicamento, Ascitis, Hepatomegalia, Aracnoides, Fosfatasa Alcalina y SGOT. Este patrón sugiere que dichos valores faltan de manera conjunta, posiblemente asociados a un mismo subconjunto de pacientes que no contaba con ciertos registros clínicos. Por otra parte, otras variables también presentan un número significativo de ausencias: Triglicéridos con 136 casos faltantes, Colesterol con 134, Cobre con 108, Plaquetas con 11 y Etapa con 6.

El patrón más frecuente corresponde a pacientes sin valores faltantes en ninguna de las variables seleccionadas, con 276 observaciones completas. El resto de los patrones muestran combinaciones variadas de ausencias, lo que refleja una estructura no aleatoria en los datos faltantes.

## 4.2 Comparación de participantes con y sin valores perdidos

Antes de realizar la imputación de datos, es útil examinar si existen diferencias relevantes entre los participantes con y sin valores perdidos en las variables seleccionadas. Esta comparación permite identificar posibles sesgos y evaluar si los datos faltantes pueden ser ignorados o si es necesario aplicar técnicas de imputación.

En este caso, se analizarán las variables Medicamento y Colesterol, que presentan una proporción significativa de valores ausentes.

```{r}
tabla_medicamento <- data_sm |>
  dplyr::select(
    Edad,
    Sexo,
    Ascitis,
    Hepatomegalia,
    Aracnoides,
    Cobre,
    Fosfatasa_Alcalina,
    SGOT,
    Albumina,
    Medicamento
  ) |>
  mutate(missing = factor(
    is.na(Medicamento),
    levels = c(FALSE, TRUE),
    labels = c("Sin valores perdidos", "Con valores perdidos")
  )) |>
  tbl_summary(
    by = missing,
    statistic = list(
      all_continuous() ~ "{mean} ({sd})",
      all_categorical() ~ "{n} ({p}%)"
    )
  ) |>
  modify_header(label = "**Variable**",
                all_stat_cols() ~ "**{level}**<br>N = {n} ({style_percent(p, digits
= 1)}%)") |>
  modify_caption("Características de los participantes según valor perdido en
**Medicamento**") |>
  bold_labels()
```

```r
tabla_colesterol <- data_sm |>
  dplyr::select(
    Edad,
    Sexo,
    Medicamento,
    Ascitis,
    Hepatomegalia,
    Aracnoides,
    Albumina,
    SGOT,
    Cobre,
    Colesterol
  ) |>
  mutate(missing = factor(
    is.na(Colesterol),
    levels = c(FALSE, TRUE),
    labels = c("Sin valores perdidos", "Con valores perdidos")
  )) |>
  tbl_summary(
    by = missing,
    statistic = list(
      all_continuous() ~ "{mean} ({sd})",
      all_categorical() ~ "{n} ({p}%)"
    )
  ) |>
  modify_header(label = "**Variable**",
                all_stat_cols() ~ "**{level}**<br>N = {n} ({style_percent(p, digits
= 1)}%)") |>
  modify_caption("Características de los participantes según valor perdido en
**Colesterol**") |>
  bold_labels()


tabla <- tbl_merge(
  tbls = list(tabla_medicamento, tabla_colesterol),
  tab_spanner = c("**Medicamento**", "**Colesterol**")
)
```

```r
{r}
tabla
```

Características de los participantes según valor perdido en **Medicamento**

| Variable | Medicamento | | Colesterol | |
|---|---|---|---|---|
| | **Sin valores perdidos** N = 312 (74.6%)[1] | **Con valores perdidos** N = 106 (25.4%)[1] | **Sin valores perdidos** N = 284 (67.9%)[1] | **Con valores perdidos** N = 134 (32.1%)[1] |
| **Edad** | 18,269 (3,865) | 19,310 (3,573) | 18,247 (3,837) | 19,139 (3,712) |
| **Sexo** | | | | |
| Hombre | 36 (12%) | 8 (7.5%) | 35 (12%) | 9 (6.7%) |
| Mujer | 276 (88%) | 98 (92%) | 249 (88%) | 125 (93%) |
| **Ascitis** | | | | |
| No | 288 (92%) | 0 (NA%) | 263 (93%) | 25 (89%) |
| Sí | 24 (7.7%) | 0 (NA%) | 21 (7.4%) | 3 (11%) |
| Unknown | 0 | 106 | 0 | 106 |
| **Hepatomegalia** | | | | |
| No | 152 (49%) | 0 (NA%) | 137 (48%) | 15 (54%) |
| Sí | 160 (51%) | 0 (NA%) | 147 (52%) | 13 (46%) |
| Unknown | 0 | 106 | 0 | 106 |
| **Aracnoides** | | | | |
| No | 222 (71%) | 0 (NA%) | 202 (71%) | 20 (71%) |
| Sí | 90 (29%) | 0 (NA%) | 82 (29%) | 8 (29%) |
| Unknown | 0 | 106 | 0 | 106 |
| **Cobre** | 98 (86) | NA (NA) | 100 (88) | 72 (58) |
| Unknown | 2 | 106 | 2 | 106 |
| **Fosfatasa_Alcalina** | 1,983 (2,140) | NA (NA) | | |
| Unknown | 0 | 106 | | |
| **SGOT** | 123 (57) | NA (NA) | 124 (57) | 109 (55) |
| Unknown | 0 | 106 | 0 | 106 |
| **Albumina** | 3.52 (0.42) | 3.43 (0.43) | 3.52 (0.40) | 3.45 (0.46) |
| **Medicamento** | | | | |
| D_penicilamina | 158 (51%) | 0 (NA%) | 140 (49%) | 18 (64%) |
| Placebo | 154 (49%) | 0 (NA%) | 144 (51%) | 10 (36%) |
| Unknown | 0 | 106 | 0 | 106 |
| **Colesterol** | | | 370 (232) | NA (NA) |
| Unknown | | | 0 | 134 |

[1] Mean (SD); n (%)

Se realizó una comparación descriptiva de las características clínicas y demográficas de los participantes según la presencia o ausencia de valores perdidos en las variables Medicamento y Colesterol.

En el caso de la variable Medicamento, el 25.4% de los participantes (n = 106) presentaron valores faltantes. Se observó que, en este grupo, la edad promedio fue ligeramente mayor de19.3 años en vez de  18.3 años Además, aunque el sexo se distribuyó de forma similar entre los grupos, todas las observaciones con valor faltante carecían también de datos en variables clínicas relacionadas como Ascitis, Hepatomegalia, Aracnoides, Fosfatasa_Alcalina, SGOT, y Cobre, lo cual sugiere un patrón de pérdida estructurado. Este hallazgo refuerza la necesidad de aplicar técnicas de imputación en lugar de eliminar casos.

Respecto a la variable Colesterol, el 32.1% de los registros (n = 134) contenían datos faltantes. Al igual que con Medicamento, se observó que los participantes con datos faltantes tendían a ser ligeramente mayores de 19.1 años contro los  18.2 años. Las variables clínicas como Albumina, SGOT y Cobre mostraron distribuciones similares entre ambos grupos, sin diferencias llamativas en las proporciones de sexo o en la distribución del tratamiento

En conjunto, estos resultados sugieren que los valores perdidos en ambas variables no son completamente aleatorios, especialmente en el caso de Medicamento, donde se evidencia un patrón conjunto de ausencias. Por tanto, el uso de imputación múltiple se justifica plenamente para evitar la pérdida de información y posibles sesgos en los análisis posteriores.

### 4.3 ¿Qué variables debo incluir en el proceso de imputación?

Debemos incluir todas las variables que serán utilizadas en los análisis posteriores, incluso aquellas que no presentan valores perdidos. Esto se debe a que el modelo de imputación debe ser tan completo como el análisis que se realizará posteriormente. De lo contrario, se podría omitir información importante para la predicción de los valores faltantes.

En el contexto del estudio de cirrosis hepática, variables como la edad, el sexo, el estado clínico, enzimas hepáticas y recuento de plaquetas aportan información clínica relevante que puede mejorar la estimación de los valores ausentes. Además, es importante asegurarse de que las variables categóricas estén correctamente codificadas como factores, para que el algoritmo `mice` pueda tratarlas adecuadamente.

```{r}
input_data <- data_sm |>
  dplyr::select(
    Edad,
    Sexo,
    Ascitis,
    Hepatomegalia,
    Aracnoides,
    Edema,
    Medicamento,
    Cobre,
    Fosfatasa_Alcalina,
    SGOT,
    Albumina,
    Colesterol,
    Trigliceridos,
    Plaquetas,
    Tiempo_Protrombina,
    Etapa
  ) |>
  mutate(
    Sexo = as.factor(Sexo),
    Ascitis = as.factor(Ascitis),
    Hepatomegalia = as.factor(Hepatomegalia),
    Aracnoides = as.factor(Aracnoides),
    Edema = as.factor(Edema),
    Medicamento = as.factor(Medicamento),
    Etapa = as.factor(Etapa)
  )
```

## 4.4 La función `mice()` para imputar datos

Para imputar datos utilizaremos la función `mice()` del paquete del mismo nombre. Entre sus argumentos, debemos especificar:

- el número de imputaciones con `m`,
- una semilla (`seed`) para que los resultados sean reproducibles, y
- el método de imputación con `method`.

En este caso, emplearemos el método `"pmm"` (predictive mean matching) para las variables numéricas continuas y `"logreg"` para variables binarias categóricas. Para las variables que no presentan valores faltantes, se asignará una cadena vacía `""`.

Primero, revisamos el orden de las variables en el conjunto `input_data`:

```{r}
names(input_data)
```

```
 [1] "Edad"              "Sexo"              "Ascitis"
"Hepatomegalia"      "Aracnoides"        "Edema"             "Medicamento"
"Cobre"
 [9] "Fosfatasa_Alcalina" "SGOT"              "Albumina"          "Colesterol"
"Trigliceridos"      "Plaquetas"         "Tiempo_Protrombina" "Etapa"
```

El método de imputación la indicaremos con el argumento `method` en el mismo orden que aparecen las variables en el dataset.

```{r}
data_imputada <- mice(
  input_data,
  m = 20,
  method = c(
    "",          # Edad
    "",          # Sexo (completo)
    "logreg",    # Ascitis
    "logreg",    # Hepatomegalia
    "logreg",    # Aracnoides
    "",          # Edema (sin NA)
    "polyreg",   # Medicamento
    "pmm",       # Cobre
    "pmm",       # Fosfatasa_Alcalina
    "pmm",       # SGOT
    "",          # Albumina (sin NA)
    "pmm",       # Colesterol
    "pmm",       # Trigliceridos
    "pmm",       # Plaquetas
    "pmm",       # Tiempo_Protrombina
    "polyreg"    # Etapa
  ),
  maxit = 20,
  seed = 123,
  print = FALSE
)
```

```r
data_imputada
```

RStudio: Notebook Output

```
Class: mids
Number of multiple imputations:  20
Imputation methods:
              Edad               Sexo             Ascitis       Hepatomegalia          Aracnoides              Edema         Medicamento
Cobre Fosfatasa_Alcalina
                ""                 ""           "logreg"           "logreg"           "logreg"                 ""           "polyreg"
"pmm"            "pmm"
              SGOT           Albumina         Colesterol       Triglicéridos  Plaquetas Tiempo_Protrombina              Etapa
             "pmm"                ""              "pmm"              "pmm"              "pmm"              "pmm"           "polyreg"
PredictorMatrix:
              Edad Sexo Ascitis Hepatomegalia Aracnoides Edema Medicamento Cobre Fosfatasa_Alcalina SGOT Albumina Colesterol Triglicéridos
Plaquetas Tiempo_Protrombina
Edad             0    1       1             1          1     1           1     1                    1    1       1          1             1
1                  1
Sexo             1    0       1             1          1     1           1     1                    1    1       1          1             1
1                  1
Ascitis          1    1       0             1          1     1           1     1                    1    1       1          1             1
1                  1
Hepatomegalia    1    1       1             0          1     1           1     1                    1    1       1          1             1
1                  1
Aracnoides       1    1       1             1          0     1           1     1                    1    1       1          1             1
1                  1
Edema            1    1       1             1          1     0           1     1                    1    1       1          1             1
1                  1
              Etapa
Edad              1
Sexo              1
Ascitis           1
Hepatomegalia     1
Aracnoides        1
Edema             1
```

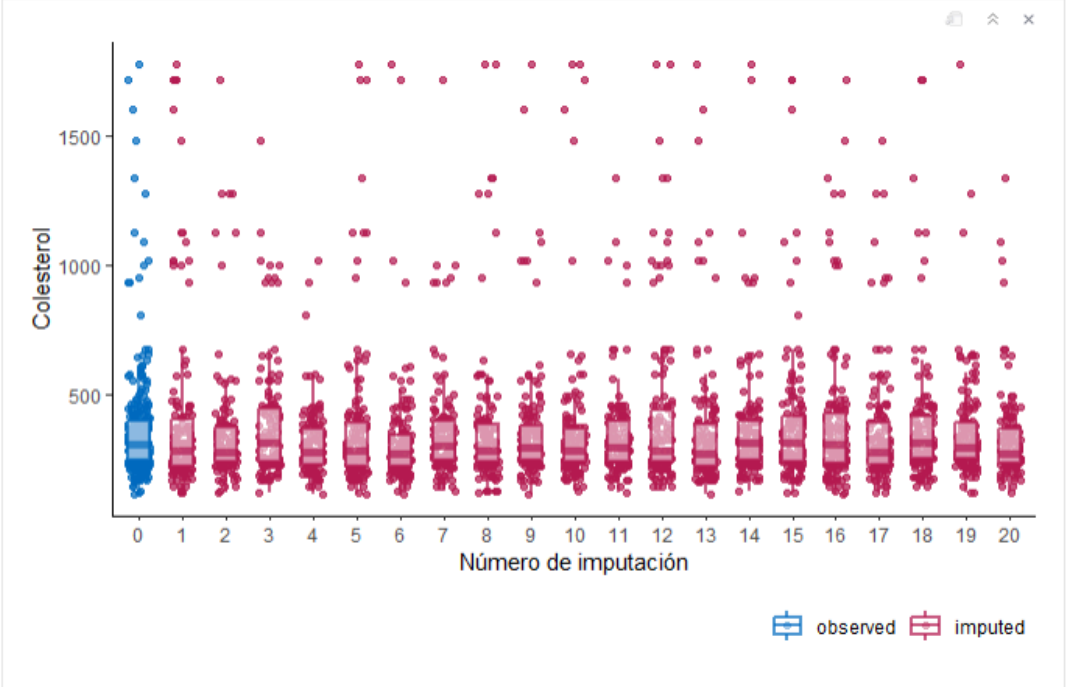
El objeto data_imputada contiene el resultado de aplicar imputación múltiple con veinte conjuntos completados. Se utilizaron métodos específicos según el tipo de variable: el método predictivo por medias (pmm) se aplicó a variables numéricas como colesterol, triglicéridos y enzimas hepáticas; el método de regresión logística (logreg) se usó para variables binarias como ascitis, hepatomegalia y aracnoides; y la regresión polinómica (polyreg) se utilizó en variables categóricas con más de dos niveles como medicamento y etapa. Las variables completas, como edad, sexo, edema y albúmina, no se imputaron y aparecen sin método asignado. La matriz de predictores indica que la mayoría de las variables se usaron como predictoras entre sí, lo que sugiere un modelo de imputación completo y coherente. Esta configuración garantiza una estimación adecuada de los valores faltantes y proporciona una base sólida para los análisis posteriores.

## 5 Analizando los datos imputados

Antes de realizar análisis posteriores sobre el dataset imputado, es fundamental explorar si los valores generados por el modelo son razonables y coherentes con los datos observados. Una forma de hacerlo es mediante gráficos de cajas y puntos, que muestran la distribución de los valores imputados frente a los valores originales en cada conjunto de imputación.

Para la variable Colesterol

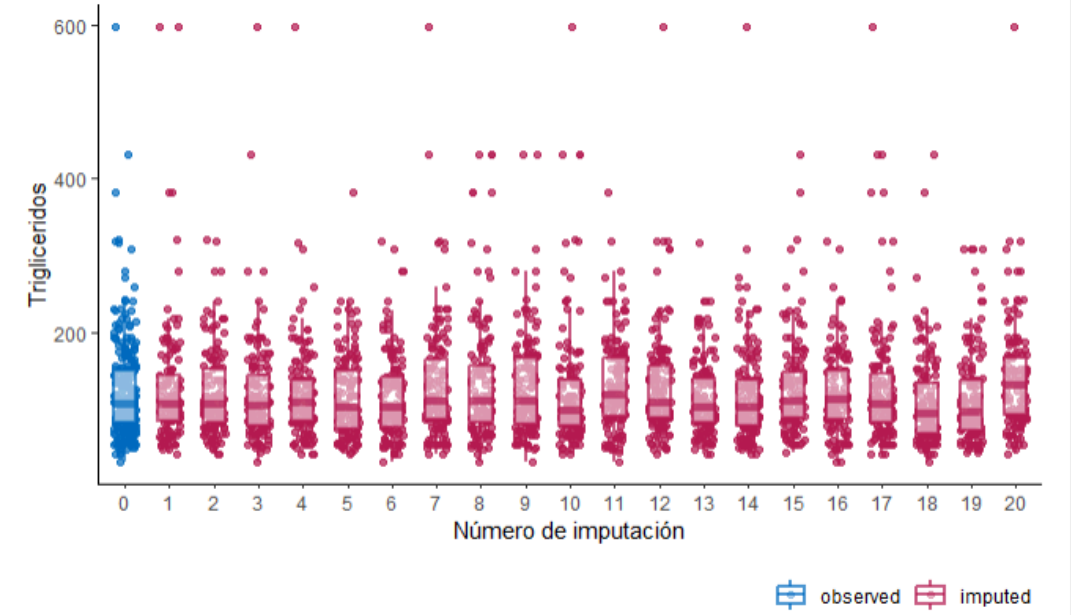
```r
{r}
ggmice(data_imputada, aes(x = .imp, y = Colesterol)) +
  geom_jitter(height = 0, width = 0.25) +
  geom_boxplot(width = 0.5, size = 1, alpha = 0.55, outlier.shape = NA) +
  labs(x = "Número de imputación", y = "Colesterol")
```




El gráfico muestra la distribución del colesterol en los datos observados (color azul) y en los veinte conjuntos imputados (color rosado). Se puede ver que los valores imputados siguen un patrón similar al de los datos originales, con una dispersión comparable y presencia de valores extremos dentro del rango esperado. Esto sugiere que el proceso de imputación fue adecuado, ya que los nuevos valores generados mantienen la estructura y variabilidad de los datos reales.

Para la variable Trigliceridos

```r
ggmice(data_imputada, aes(x = .imp, y = Trigliceridos)) +
  geom_jitter(height = 0, width = 0.25) +
  geom_boxplot(width = 0.5, size = 1, alpha = 0.55, outlier.shape = NA) +
  labs(x = "Número de imputación", y = "Trigliceridos")
```




El gráfico muestra cómo se comportan los valores de triglicéridos después de la imputación. A la izquierda, en azul, están los valores reales observados, y a la derecha, en rosa, los generados por el modelo en las veinte imputaciones. Se puede notar que las distribuciones son muy similares: los valores imputados se concentran en el mismo rango que los datos originales, con una variabilidad comparable y sin desviaciones evidentes. Esto indica que la imputación fue exitosa y que los valores faltantes fueron completados de forma coherente con la información real del estudio

## Comparando la distribución de variables categóricas imputadas

Para las variables categóricas, es útil evaluar si la distribución de los valores imputados es similar a la de los valores observados. Esto permite verificar si el modelo de imputación mantuvo una proporción razonable entre las categorías.

A continuación, se transforma el objeto `data_imputada` a formato largo y se construye una tabla de proporciones para la variable `Medicamento`.undefined**undefined**undefined

```{r}
# Convertir el objeto imputado a formato "long"
data_imputada_l <- complete(data_imputada, "long", include = TRUE)
```

```{r}
# Marcar si el valor es imputado o observado
data_imputada_l <- data_imputada_l %>%
  mutate(imputed = .imp > 0,
         imputed = factor(imputed,
                          levels = c(FALSE, TRUE),
                          labels = c("Observado", "Imputado")))
# Tabla de proporciones para la variable Medicamento
prop.table(
  table(data_imputada_l$Medicamento, data_imputada_l$imputed),
  margin = 2
)
```

```
                Observado  Imputado
  D_penicilamina 0.5064103 0.5160287
  Placebo        0.4935897 0.4839713
```

La tabla de proporciones muestra que la distribución de la variable Medicamento se mantuvo muy similar entre los datos observados e imputados. En los casos observados, el 50.6 % de los pacientes recibieron D-penicilamina y el 49.4 % recibieron placebo. En los datos imputados, estas proporciones fueron 51.6 % y 48.4 % respectivamente. La diferencia entre ambas distribuciones es mínima, lo que indica que el modelo de imputación respetó el patrón original de la variable y generó valores coherentes. Esto refuerza la validez de los resultados imputados para esta variable categórica.

## 5.1 Procedimientos adicionales luego de la imputación

Una vez imputados los datos, se pueden realizar análisis inferenciales con total normalidad. En el caso de regresión logística, basta con aplicar la función with() directamente sobre el objeto data_imputada, y si se desea combinar los resultados de múltiples imputaciones, se puede usar pool(). Si se utiliza el paquete gtsummary, este realiza internamente la agrupación de imputaciones, por lo que solo se necesita aplicar with().

A continuación se presenta un ejemplo de regresión logística multivariada utilizando datos imputados, donde se modela la probabilidad de encontrarse en una etapa clínica más avanzada en función de variables como edad, sexo, enzimas hepáticas, lípidos y otras características clínicas.

```{r}
tabla_multi <-
  data_imputada |>
  with(glm(Etapa ~ Edad + Sexo + Ascitis + Hepatomegalia + SGOT +
             Fosfatasa_Alcalina + Colesterol + Trigliceridos + Albumina,
             family = binomial(link = "logit"))) |>
  tbl_regression(
    exponentiate = TRUE,
    label = list(
      Sexo ~ "Sexo",
      Ascitis ~ "Ascitis",
      Hepatomegalia ~ "Hepatomegalia",
      SGOT ~ "SGOT (U/L)",
      Fosfatasa_Alcalina ~ "Fosfatasa Alcalina (U/L)",
      Colesterol ~ "Colesterol (mg/dL)",
      Trigliceridos ~ "Triglicéridos (mg/dL)",
      Albumina ~ "Albúmina (g/dL)",
      Edad ~ "Edad (años)"
    )
  ) |>
  bold_p(t = 0.05) |>
  modify_header(estimate = "**OR ajustado**", p.value = "**p valor**")
```

| Characteristic | OR ajustado | 95% CI | p valor |
|---|---|---|---|
| Edad (años) | 1.00 | 1.00, 1.00 | **0.035** |
| Sexo | | | |
| Hombre | — | — | |
| Mujer | 4.85 | 0.82, 28.8 | 0.082 |
| Ascitis | | | |
| No | — | — | |
| Sí | 8,179 | 0.00, Inf | >0.9 |
| Hepatomegalia | | | |
| No | — | — | |
| Sí | 5,750,405 | 0.00, Inf | >0.9 |
| SGOT (U/L) | 1.02 | 1.00, 1.04 | **0.015** |
| Fosfatasa Alcalina (U/L) | 1.00 | 1.00, 1.00 | 0.5 |
| Colesterol (mg/dL) | 1.00 | 0.99, 1.01 | 0.7 |
| Triglicéridos (mg/dL) | 1.01 | 1.0, 1.03 | 0.15 |
| Albúmina (g/dL) | 0.87 | 0.20, 3.79 | 0.9 |

Abbreviations: CI = Confidence Interval, OR = Odds Ratio

El modelo que se usó ayuda a entender qué características de los pacientes están relacionadas con tener una etapa más avanzada de cirrosis. De todas las variables analizadas, dos mostraron una relación importante: la edad y la enzima SGOT.

Aunque el valor del odds ratio para la edad fue 1.00, el valor p fue significativo (0.035), lo que indica que a mayor edad, aumenta ligeramente la probabilidad de estar en una etapa más avanzada de la enfermedad.

La enzima SGOT también fue significativa. Los pacientes con niveles más altos de SGOT tienen más probabilidad de estar en una etapa avanzada, ya que el valor p fue 0.015, lo cual respalda esta relación.

Las demás variables, como el sexo, la presencia de ascitis, hepatomegalia, colesterol, triglicéridos y albúmina, no mostraron una relación clara con la etapa de la enfermedad. Algunas de ellas dieron resultados muy extremos o inestables, por lo que no se puede sacar una conclusión clara en este análisis.

Los resultados muestran que la edad y la SGOT son factores importantes que podrían estar relacionados con el avance de la cirrosis, mientras que el resto no parece tener un peso importante en este modelo. Además, el uso de imputación ayudó a completar los datos y hacer un análisis más sólido.